

SVKM's
D. J. Sanghvi College of Engineering

**Program: B.Tech in Information
Technology**

Academic Year: 2022

Duration: 3 hours

Date: 14.01.2023

Time: 10:30 am to 01:30 pm

Subject: Statistical Analysis (Semester V)

Marks: 75

Instructions: Candidates should read carefully the instructions printed on the question paper and on the cover page of the Answer Book, which is provided for their use.

- (1) This question paper contains two pages.
- (2) **All Questions are Compulsory.**
- (3) All questions carry equal marks.
- (4) **Answer to each new question is to be started on a fresh page.**
- (5) **Figures in the brackets on the right indicate full marks.**
- (6) **Assume suitable data wherever required, but justify it.**
- (7) Draw neat, labeled diagrams, wherever necessary.

Question No.		Max. Marks																	
Q1 (a)	<div><div>i. Write the Basic Steps of the Research Process.</div><div>ii. Identify the sampling technique from the given scenario:<div><div>a. An airline company wants to survey its customers one day, so they randomly select 15 flights that day and survey every passenger on those flights.</div><div>b. All employees of the company are listed in alphabetical order. From the first 10 numbers, you randomly select a starting point: number 6. From number 6 onwards, every 10th person on the list is selected (6, 16, 26, 36, and so on), and you end up with a sample of 100 people.</div><div>c. The company has 800 female employees and 200 male employees. You want to ensure that the sample reflects the gender balance of the company, so you sort the population into two strata based on gender. Then you use random sampling on each group, selecting 80 women and 20 men, which gives you a representative sample of 100 people.</div></div></div></div>	<div>[04]</div> <div>[03]</div>																	
Q1 (b)	<div>A food services manager for a baseball park wants to know if there is a relationship between gender and the preferred condiment on a hot dog. The following table summarizes the results. Test the hypothesis with a significance level of 10%.</div> <table><tr><th rowspan="4">Gender</th><th colspan="4">Condiment</th></tr><tr><th></th><th>Ketchup</th><th>Mustard</th><th>Relish</th></tr><tr><td>Male</td><td>15</td><td>23</td><td>10</td></tr><tr><td>Female</td><td>25</td><td>19</td><td>08</td></tr></table>	Gender	Condiment					Ketchup	Mustard	Relish	Male	15	23	10	Female	25	19	08	[08]
Gender	Condiment																		
			Ketchup	Mustard	Relish														
	Male		15	23	10														
	Female	25	19	08															

Q2 (a)	<p>A recent large survey of a random sample of Australian children asked about weekly hours of internet use in three age groups. The following table shows the mean and standard deviation of the number of hours of internet use per week and the total number of children surveyed for each age group. Calculate an approximate 95% confidence interval for the mean number of hours of internet use per week in each group.</p> <table><tr><th colspan="4">Internet Use</th></tr><tr><th>Age Group (years)</th><th>Mean (hours/week)</th><th>Standard deviation</th><th>Number surveyed</th></tr><tr><td>5–8</td><td>3.29</td><td>4.29</td><td>2150</td></tr><tr><td>9–11</td><td>5.75</td><td>6.17</td><td>2530</td></tr><tr><td>12–14</td><td>9.95</td><td>7.81</td><td>1250</td></tr></table> <p style="text-align: center;">OR</p> <p>Suppose a researcher has access to a list of all pilots who are members of the Air Line Pilots Association. If this list is used as a frame for the study, she can randomly select a sample of pilots, contact them, and ascertain their ages. From 89 of these pilots so selected, she learns that 48 are more than 40 years of age. Construct an 85% confidence interval to estimate the population proportion of commercial airline pilots who are more than 40 years of age. What is the higher value of this interval?</p>	Internet Use				Age Group (years)	Mean (hours/week)	Standard deviation	Number surveyed	5–8	3.29	4.29	2150	9–11	5.75	6.17	2530	12–14	9.95	7.81	1250	[08]
Internet Use																						
Age Group (years)	Mean (hours/week)	Standard deviation	Number surveyed																			
5–8	3.29	4.29	2150																			
9–11	5.75	6.17	2530																			
12–14	9.95	7.81	1250																			
Q2 (b)	<p>1000 students at a college were graded according to their I.Q. and the economic condition of their homes. Find out whether there is any association between economic condition at home and I.Q. using goodness of fit test at 5% level of significance.</p> <table><tr><th rowspan="2">Economic Condition</th><th colspan="2">I.Q.</th></tr><tr><th>High</th><th>Low</th></tr><tr><td>Rich</td><td>460</td><td>140</td></tr><tr><td>Poor</td><td>240</td><td>160</td></tr></table>	Economic Condition	I.Q.		High	Low	Rich	460	140	Poor	240	160	[07]									
Economic Condition	I.Q.																					
	High	Low																				
Rich	460	140																				
Poor	240	160																				
Q3 (a)	<p>For the following data</p> <table><tr><td>Marks</td><td>0-10</td><td>10-20</td><td>20-30</td><td>30-40</td><td>40-50</td><td>50-60</td><td>60-70</td></tr><tr><td>No.of Students</td><td>5</td><td>10</td><td>18</td><td>30</td><td>20</td><td>12</td><td>5</td></tr></table> <p>i. Calculate the mean. ii. Calculate the median and mode for the data.</p> <p style="text-align: center;">OR</p> <p>Consider the dataset given below 11,23,32,26,16,19,30,14,16,10 Generate the five number summary plot.</p>	Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	No.of Students	5	10	18	30	20	12	5	<p>[02] [03]</p> <p>[05]</p>				
Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70															
No.of Students	5	10	18	30	20	12	5															

Q3 (b)	<p>Four brands of flashlight batteries are to be compared by testing each brand in five flashlights. Twenty flashlights are randomly selected and divided randomly into four groups of five flashlights each. Then each group of flashlights uses a different brand of battery. The lifetimes of the batteries, to the nearest hour, are as follows.</p> <table><tr><th>Brand A</th><th>Brand B</th><th>Brand C</th><th>Brand D</th></tr><tr><td>42</td><td>28</td><td>24</td><td>20</td></tr><tr><td>30</td><td>36</td><td>36</td><td>32</td></tr><tr><td>39</td><td>31</td><td>28</td><td>38</td></tr><tr><td>28</td><td>32</td><td>28</td><td>28</td></tr><tr><td>29</td><td>27</td><td>33</td><td>25</td></tr></table> <p>Preliminary data analyses indicate that the independent samples come from normal populations with equal standard deviations. At the 5% significance level, use ANOVA to determine whether there is a significant difference in mean lifetime among the four brands of batteries?</p>	Brand A	Brand B	Brand C	Brand D	42	28	24	20	30	36	36	32	39	31	28	38	28	32	28	28	29	27	33	25	[10]
Brand A	Brand B	Brand C	Brand D																							
42	28	24	20																							
30	36	36	32																							
39	31	28	38																							
28	32	28	28																							
29	27	33	25																							
Q4 (a)	<p>A random sample of size 16 has 53 as the mean. The sum of squares of deviations taken from mean is 135. Can this sample be regarded as taken from the population having 56 as the mean? Obtain 95% and 99% confidence limits of the mean of the population.</p> <p style="text-align: center;">OR</p> <p>The mean population of a random sample of 400 villages in Jaipur district was found to be 400 with a standard deviation of 12. The mean population of a random sample of 400 villages in Meerut district was found to be 395 with a standard deviation of 15. Is the difference between the two means statistically significant? 1% significance level.</p>	[08]																								
Q4 (b)	<p>Discuss Ridge Regression and Lasso Regression in detail.</p> <p style="text-align: center;">OR</p> <p>Discuss multicollinearity with the help of a suitable example.</p>	[07]																								
Q5 (a)	<p>Obtain regression equation of Y on X and estimate Y when X=55 from the following.</p> <table><tr><td>X</td><td>40</td><td>50</td><td>38</td><td>60</td><td>65</td><td>50</td><td>35</td></tr><tr><td>Y</td><td>38</td><td>60</td><td>55</td><td>70</td><td>60</td><td>48</td><td>30</td></tr></table>	X	40	50	38	60	65	50	35	Y	38	60	55	70	60	48	30	[07]								
X	40	50	38	60	65	50	35																			
Y	38	60	55	70	60	48	30																			
Q5 (b)	<p>Solve any two.</p> <p>i. Explain Types of errors in Testing of Hypothesis with suitable example.</p> <p>ii. Demonstrate Two-tailed and One-tailed tests of hypothesis with neat diagram.</p> <p>iii. Describe the procedure of Hypothesis testing with an example.</p> <p>iv. Describe the statistical test to determine whether two population means are different when the variances are known and the sample size is large</p>	<p>[04]</p> <p>[04]</p> <p>[04]</p> <p>[04]</p>																								